

TRENDS IN DEALING WITH TERM BASES

Attila IMRE, Ph.D.*

“Petru Maior” University of Tîrgu Mureş

Abstract

It is common knowledge that the turn of the century faced dramatic changes in technology. Once we accept that we can speak about a revolution of technology, it is obvious that it has had an invigorating effect upon all fields, including translation studies. As a result, the revolution of translation led from “paper and rubber assisted translation” (PRAT) to computer-assisted translation (CAT) in the 21st century, as Daniel Gouadec correctly observed it in 2007.

Furthermore, today there is no more translation as such, but we have instead translation projects full with all cutting edge technology: machine translation (MT), translation memories (TM) and term bases (TB). The present article tries to summarize a POSDRU project lasting for 18 months, a period during which we were focusing on creating a TB based on the published dictionaries in Romania over a period of fifteen years (1999–2014). The starting point for the research was the fact that in order to remain competitive on the translation market, the present day professional translator needs an extended set of competences, particularly focusing on technical skills. The debate whether MT will take over human translators has somehow lost importance, due to the emergence of CAT-tools, which simply ‘embedded’ MT as an add-in option, which can be set active or not. Thus CAT-tools really assist human translators in creating better quality translations ‘here and now’, in case translators are familiar with the possibilities of importing–exporting TMs, TBs, etc.

We are primarily concerned with TBs, as terminology research for translators is vital in providing high quality translations that MT will never be able to achieve. Although our initial claim was that printed dictionaries offer by far a better quality than unreliable online sources (dictionaries and term bases), we tend to rephrase that claim. On the one hand, printed sources contain many errors, presented in a couple of articles, whereas the combination of online sources seems to surpass in number all the entries from the printed dictionaries.

It is also important to revise another deeply rooted concept. In our case, we started from Romanian–English, English–Romanian dictionaries of legal terms, but we soon discovered that a clear-cut set of terms is not possible. Legal terms are often combined with economics, geography, IT, history, etc., and the major advantage of online term bases derives from this fact: these TBs do not promise a legal or economical list of entries, they offer them as is. Furthermore, their set of entries is not set in stone, it can grow day by day, and serious translation errors are fewer and fewer.

Professional translators may be reluctant to use them, but the trend is obvious: better and better quality online TBs will take over, and human translators should learn the way how to use and embed them in CAT-tools to their benefit.

Keywords: machine translation, computer-assisted translation, term bases, quality assurance, Romanian–English, legal dictionaries.

1. Introduction

The dramatic changes towards the end of the 20th century resulted – among others – in the revolution of technology, and the spread of computers led to the existence of a “global village” envisaged by McLuhan (1962, p. 21), even if he develops the concept from the point of view of the media. Interestingly, after the Second World War research on machine translation (MT) also started, although somewhat delayed due to the infamous ALPAC report in 1966 (Hutchins, 2003). Although there were many allegations that MT would eventually replace human translation (cf. Imre, 2013, pp. 223–230), this is still thriving, partially due to the emergence of computer-assisted translation (CAT) tools. Daniel Gouadec predicted (2007, p. 109) that “paper and rubber assisted translation” (PRAT) was about to disappear, and it would be replaced by CAT, a statement we cannot deny in the 21st century.

As a result, there is no more translation as such, but we have translation projects instead, packed with all cutting edge technology, under the hood of CAT, which encapsulated MT, translation memories (TM) and term bases (TB) as well. The basic idea of CAT is “never translate the same thing twice”, meaning that once we have a translation of a sentence (technically speaking: segment), the translation is preserved in a database, which can be reused when a similar text is translated next time.

As translators specialize in various fields, it is obvious that they must have a collection of specialized terms, which may be preserved within a CAT tool (.csv format), or separately, for instance in an Excel file (.xls or .xlsx format). The larger the database, the more matches translators will have during the next translation project, which contributes directly to quality assurance combined with shorter processing / editing time, exemplified in the list below:

Romanian	English
<i>abandon</i>	<i>abandon</i>
<i>abandon</i>	<i>abandonment</i>
<i>abandon</i>	<i>dereliction</i>
<i>abandon</i>	<i>desertion</i>
<i>abandon</i>	<i>renunciation</i>

Table 12. Terms for the Romanian ‘abandon’

Table 1 shows the possible English translations for the Romanian *abandon*, which may be easily found either in printed dictionaries or online databases (online dictionaries or term bases). A professional translator specialized in legal translations will soon realize that this entry is important in legal texts, thus it is vital for a TB to be modified: new entries should be added, older ones completed or even deleted when they are not suitable any more. But this already leads us to the next section.

2. The project

* Corresponding author email: amireittal@gmail.com.

Seeing the obvious (r)evolution of translation industry, we consider it important to have reliable collections of specialized terms in various (major) fields, such as technology, medicine or law. We applied successfully for a POSDRU postdoctoral research in 2014, setting the goal of creating a Romanian–English legal term base, relying on the bilingual dictionaries published in Romania between 1999 and 2014. Professional translators on ProZ.com often complain about the poor quality of existing Romanian–English legal term bases, and it seemed a reasonable choice to start with.

During the project we were able to browse through more than a dozen dictionaries, and checked about 330,000 entries. As a result, we could create a collection of around 50,000 entries, and in the case of the Romanian *abandon*, we can offer further possible translations: *cession* (Lister & Veth, 2010, p. 315; Hanga & Calciu, 2009, p. 7; Botezat, 2011, p. 3; Voroniuc, 2011, p. 203), *surrender* (Dumitrescu, 2009, p. 3) or *waiver* (Lozinschi, 2008, p. 1), but it is obvious that further entries for the Romanian term are necessary.

First of all, expressions containing *abandon* are needed: *abandon de domiciliu* (abandonment of domicile), *abandon de familie* (family abandonment), *abandon de mărfuri* (abandonment of cargo), *declarație de abandon* (notice of abandonment), *abandonul vasului scufundat* (abandonment of a sunken ship), *abandon definitiv al domiciliului conjugal* (obstinate desertion – US), *abandon prin efectul legii* (surrender by act and operation of the law), terms which are found in the previously mentioned dictionaries. Although some of them belong to commercial terms. What is more, the Romanian term must be completed with a synonymous noun (*abandonare*) with somewhat different translations and expressions. To make matters worse, there is also a verb (*a abandona*) associated with the noun, which results in further entries. Thus the set is completed with new entries, such as *discontinuance* or *nolle prosequi*, hoping that during translation at least one match will be suitable. We created a TB with the Romanian *abandon* and created a project in *memoQ* CAT tool, which is an integrated translation environment, being able to handle MTs, TMs and TBs. Figure 1 below shows how these entries come up during translating a segment containing *abandon*:

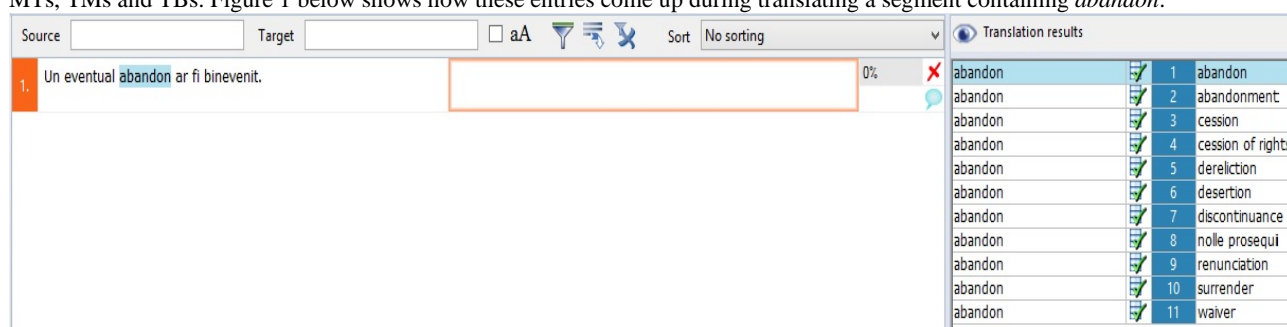


Figure 30. Matches for *abandon* in *memoQ*

Figure 1 presents a translation project in *memoQ*, when only one segment should be translated, containing *abandon*. The right side shows the *Translation results*, eleven in number, out of which one is supposed to be the correct one, and an expert translator can obviously make the right decision based on the context. In the long run, the database will grow, thus better and better matches will be shown, offering another advantage: one term will always be translated the same way, if needed (in the case of specialized translations). Yet, a question arises, namely whether we need TBs or we can handle translations in the traditional way.

3. Trends in dealing with term bases

There are many articles and books written regarding the competences of the 21st century translator, and their technical skills are highlighted (e.g. Bowker, 2002; Anderman & Rogers, 2003; Kis & Mohácsi-Gorove, 2008), without which a translator will disappear from the translation market. It is also interesting to observe that the debate whether MT will take over or not human translators has somehow lost importance, as CAT-tools use MT variants as an ‘embedded’ option, which can be set active or completely ignored. For instance, *memoQ* can activate eleven MT tools, some of which are illustrated below:



Figure 31. MT in *memoQ*

More than that, there are companies hiring professional translators who translate in a cloud-based system and their translations are stored in secured company storage systems in the form of TMs and TBs. The IATE database (<http://iate.europa.eu/SearchByQueryLoad.do?method=load>), which is the EU's multilingual term base, is a fortunate exception to this company rule, but *Quest* is only accessible for employees. CAT tools having access to large collections of TMs and TBs will produce high quality matches, fulfilling the promise: they really *assist* human translators in creating better quality translations ‘here and now’. The two conditions for the translator are to have access to these databases and to be familiar with the possibilities of importing–exporting TMs and TBs. As more and more freely accessible TBs emerge, we are primarily concerned with TBs, as terminology research for translators is vital in providing high quality translations.

Although our initial claim was that printed dictionaries offer by far a better quality than unreliable online sources (dictionaries and term bases), we tend to rephrase that claim. On the one hand, printed sources contain many errors, presented in a couple of articles (Imre, 2014a; Imre, 2014b), whereas on the other hand the combination of online sources seem to surpass in number of entries all printed dictionaries taken altogether (<https://glosbe.com/>, <http://ro-en.gsp.ro/>, <http://hallo.ro/>, <http://www.activ-traduceri.ro/>, or <http://www.dictionarromanenglez.ro/>).

At this stage, it is important to revise another deeply rooted concept. In our case, we started from Romanian–English, English–Romanian dictionaries of legal terms, but we soon discovered that a clear-cut set of terms is not possible. Legal terms are often combined with economics, geography, IT, history, etc., and the major advantage of online term bases derives from this fact: these TBs do not promise a *legal* or *economical* list of entries, they offer them *as is*. Furthermore, their set of entries is not set in stone, it can grow day by day, and serious translation errors are fewer and fewer. Hence, we predict that the common distrust towards online resources will soon evaporate, as they are constantly better. Even *Google Translate* – one of the leading MT sources – is getting better and better. For the time being professional translators may feel reluctant to use them, but the trend is obvious: online TBs are constantly fed by previous translations carried out by human translators, and these search and match machines are capable of finding parallel corpora in multiple languages (cf. the EU's database). The CAT tool we are familiar with (*memoQ*) has a *LiveDocs* function, which can pair two parallel texts and reuse them in offering matches for completely new segments that are connected to this pair of texts.

4. Conclusions

It is worth mentioning that policies regarding TBs change. Although there is a trend for specialized companies to try to protect them, TBs become freely available online. Search engines may be used for the benefit of everybody, and TBs will offer us more and more reliable matches based on probability (number of occurrences). Professional translators can make the difference between poor results and quality ones, but we have to agree that improvements are fast. A few years ago a segment was interpreted from full-stop to full-stop, which meant that all abbreviations, date sections (day, month, year) counted as separate segments; today CAT tools can interpret both dates and abbreviations, thus segment boundaries are more realistic. Furthermore, first and last names are hardly translated (e.g. *Frunză* tends to be preserved instead of *Leaf*), and even longer combinations of words (the majority of specific terms) have correct variants.

Certain refinements are still needed. For instance, best results are obtained only when one of the languages involved (source or target) is English, and specific fields seem to be better covered (e.g. medicine, probably due to common Latin terms and its importance in human life).

We also estimate that free online TBs will soon overwhelm us, and a further challenge of translators will be the selection of proper matches out of a bulk of hits. After all, the cycle of struggle for better and better quality in translations should be kept up combined with ever shorter deadlines. Once we even lacked basic writing and translation tools, then half a decade ago automated translation emerged. Today we handle thousands of pages of texts in unified files, extracting relevant matches out of them, hoping for improved quality. McLuhan's global village brought about a McWorld and the McLanguage of American English (Snell-Hornby, 2006, p. 132), which is not a menace to all other languages – partially, at least – due to professional translators striving for proper translation of terms and their CAT tools capable of preserving terms and segments in TBs and TMs. If our approach is correct, then human translators should have no reservations about modern translation tools and should focus on handling them effectively to their benefit.

Acknowledgements

The research presented in this paper was supported by the European Social Fund under the responsibility of the Managing Authority for the Sectoral Operational Programme for Human Resources Development (*Sistem integrat de îmbunătățire a calității cercetării doctorale și postdoctorale din România și de promovare a rolului științei în societate*), as part of the grant POSDRU/159/1.5/S/133652.

References

- Anderman, G. M., & Rogers, M. (Eds.). (2003). *Translation Today: Trends and Perspectives*. Multilingual Matters.
- Botezat, O. (2011). *Dicționar juridic român-englez / englez-român* (2nd ed.). București: C.H. Beck.
- Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa Press.
- Dumitrescu, D. (2009). *Dicționar juridic român-englez*. București: Akademos Art.
- Gouadec, D. (2007). *Translation as a Profession*. John Benjamins Publishing.
- Hanga, V., & Calciu, R. (2009). *Dicționar juridic englez-român și român-englez*. București: Lumina Lex.
- Hutchins, J. (2003). ALPAC: The (In)Famous Report. In S. Nirenburg, H. L. Somers, & Y. A. Wilks (Eds.), *Readings in Machine Translation* (A Bradford Book, pp. 131–136). Cambridge, MA; London, England: The MIT Press.
- Imre, A. (2013). *Traps of Translation*. Brașov: Editura Universității “Transilvania.”
- Imre, A. (2014a). Difficulties of creating a legal term base. In I. Boldea (Ed.), *Studies on Literature, Discourse and Multicultural Dialogue* (pp. 276–284). Târgu-Mureș: Arhipelag XXI Press.
- Imre, A. (2014b). Ways to enhance legal dictionaries. *Communication, Context, Interdisciplinarity*, (3), 519–527.
- Kis, B., & Mohácsi-Gorove, A. (2008). *A fordító számítógépe*. Bicske: Szak Kiadó.
- Lister, R., & Veth, K. (2010). *Dicționar juridic englez-român / român-englez*. (R. Dinulescu, Trans.). București: Niculescu.

Lozinschi, S. (2008). *Dicționar juridic Român - Englez*. București: Editura Smaranda.
McLuhan, M. (1962). *The Gutenberg Galaxy: The Making of Typographic Man*. University of Toronto Press.
Snell-Hornby, M. (2006). *The Turns of Translation Studies: New Paradigms Or Shifting Viewpoints?* John Benjamins Publishing.
Voroniuc, A. (2011). *Dicționar englez-român / român-englez de termeni economici și juridici*. Iași: Polirom.

Attila Imre (2015) Trends in dealing with term bases, in Popa, Alina & Ioncioaia, Florea (eds)
Proceedings of the 2nd CommScie International Conference "Challenges for Sciences and
Society in Digital Era", Iași: Editura Pim, pag. 199-202. ISBN: 978-606-13-2892-5.